

Contando e Detectando Mentiras: Efeito do *Feedback* sobre o Desempenho

Nicolau Chaud de Castro Quinta
Cristiano Coelho¹
Universidade Católica de Goiás

RESUMO - Estudos nos quais pessoas são testadas ao detectar mentiras mostram que a maioria da população parece ter essa habilidade pouco desenvolvida. O objetivo do presente trabalho foi avaliar o efeito do *feedback* dado após cada julgamento sobre a detecção de mentiras, com exposição prolongada à situação experimental, na qual detector e emissor interagiam frente a frente. Os resultados mostraram que o *feedback* levou a uma melhora do desempenho de todos os detectores, com percentagens de acerto de até 100% em uma sessão. Contudo, o desempenho dos detectores mostrou pouca estabilidade e os dados sobre a generalização para outros emissores foram inconclusivos. Medidas independentes do comportamento dos emissores não revelaram diferenças consistentes entre verdades e mentiras, ainda que relatos pós-experimentais tenham apontado nessa direção. Foi discutida a importância de uma análise de dados individualizada e a necessidade de delineamentos que isolem a aprendizagem de detector e emissor.

Palavras-chave: detecção; mentira; treino; *feedback*.

Telling and Detecting Lies: The Effect of Feedback on Performance

ABSTRACT - Studies in which people are tested for detecting lies have shown that most of the population seems to have poorly developed lie-detection skills. The goal of this study was to assess the effect of feedback given after each judgment on lie detection, with extended exposure to an experimental situation in which detector and sender interacted face-to-face. Results showed that feedback led to performance improvement for all detectors, with accuracy rates up to 100% in a single session. However, detectors' performance showed little stability, and the data about generalization to other senders were inconclusive. Independent measures of senders' behavior didn't reveal consistent differences between truths and lies, even though post-experimental reports have pointed in this direction. The importance of an individualized data analysis and the need for experimental designs that isolate learning of detector and sender were discussed.

Keywords: detection; lie; training; feedback.

A mentira faz parte de nossas vidas. Nossas relações sociais habituais envolvem mentir e ouvir mentiras, mesmo que nem sempre (ou quase nunca) nos atentemos a isso. Estudos naturalísticos sobre a mentira mostram que, em média, as pessoas contam uma mentira a cada três interações sociais (DePaulo & Kashy, 1998). Algumas dessas mentiras são brandas e têm funções de regulação social, enquanto outras podem ser mais sérias e prejudicar seu alvo em vantagem daquele que conta a mentira (DePaulo, Kashy, Kirkendol, Wyer & Epstein, 1996). Mas como saber quando uma pessoa está mentindo? Cientistas de diversas áreas do conhecimento têm se ocupado dessa pergunta, e os resultados de algumas décadas de pesquisa mostram que não existe uma única resposta para ela (Ekman, 2001; Iacono & Lykken, 2002).

O termo *detector de mentiras* tem sido tradicionalmente associado ao teste do polígrafo, que consiste em um conjunto de aparelhos ligados ao corpo de uma pessoa que registram diversas medidas fisiológicas (como respiração, frequência cardíaca, pressão sanguínea, sudorese, resposta eletrodérmica etc.). A premissa básica desse teste é que ocorrem alterações neste conjunto de medidas enquanto uma pessoa mente;

até o ponto em que essas alterações são de fato correlatas à mentira e o teste é capaz de detectá-las, ele será um detector de mentiras eficiente. Resultados agregados de estudos de laboratório e de campo revelam uma acurácia média do teste em torno de 85% (Ekman, 2001).

A detecção de mentiras também vem sendo estudada pela Psicologia, mas de acordo com uma outra perspectiva metodológica: por meio da observação a olho nu do comportamento da pessoa que mente. De forma semelhante, até o ponto em que ocorrem mudanças externamente observáveis no comportamento de uma pessoa enquanto ela conta uma mentira, e um observador seja capaz de discriminá-las, este será um detector de mentiras eficiente.

Essas mudanças comportamentais recebem o nome de *sinais de mentira*. Os sinais de mentira não podem ser descritos como classes de respostas absolutas, pois sua definição é necessariamente relacional. Um determinado comportamento (como, por exemplo, inclinar a cabeça) só será um sinal de mentira se ocorrer regularmente com mais (ou menos) frequência quando a pessoa mente em relação a quando fala a verdade. Da mesma forma, sinais de mentira podem ser propriedades de comportamentos, como elevação no tom de voz, aumento no ritmo da fala etc., desde que tais mudanças sejam regulares ao diferenciar a fala sincera da mentirosa. A detecção de mentiras não será, assim, o reconhecimento

¹ Endereço: Laboratório de Análise Experimental do Comportamento – Universidade Católica de Goiás. Av. Universitária 1.440, Setor Universitário. Goiânia, GO. CEP 74605-010. E-mail: ncquinta@terra.com.br.

do comportamento de mentir em si, mas dessas mudanças concomitantes no comportamento do mentiroso.

Os dados da literatura em detecção de mentiras por meio da observação do comportamento do mentiroso são mais pessimistas: a acurácia dos observadores nas tarefas de detecção raramente passa de 60%, sendo 50% considerado acaso ou chance (DePaulo, Kirkendol, Tang & O'Brien, 1988).

Em um estudo típico (DePaulo & cols., 1988), estudantes universitários foram filmados enquanto relatavam suas opiniões a respeito de tópicos controversos. Essas pessoas haviam sido previamente avaliadas a respeito de suas posições dentro de tais tópicos, e posteriormente instruídas a fazerem relatos honestos ou mentirosos para cada uma de suas opiniões. Foram feitas composições de fitas de vídeo contendo todos os relatos de cada um dos participantes (com cada fita contendo entre 32 e 56 relatos), e essas fitas foram mostradas a um conjunto de observadores, que deveriam julgar o grau de honestidade de cada relato. Nesse trabalho, foram manipuladas as informações dadas aos emissores (as pessoas que faziam os relatos) a respeito do número de canais de seu comportamento que estariam sendo registrados (áudio, imagem, áudio e imagem, imagem só da face, imagem só do corpo, imagem de corpo e face etc.), de modo que seriam estes os canais de comunicação que eles deveriam controlar caso quisessem evitar revelar a mentira. A hipótese dos autores era de que quanto maior o número de canais registrados, mais difícil seria para o emissor manter tal controle, de forma que a mentira seria mais facilmente detectada. Os dados mostraram uma fraca evidência a favor dessa hipótese, e apenas quando o emissor deveria controlar todos os canais. De forma geral, o desempenho dos detectores não foi melhor que aquele esperado ao acaso.

Em um outro estudo, Ekman e O'Sullivan (1991) trabalharam com 10 estudantes do curso de Enfermagem que deveriam assistir a vídeos em uma tela e relatar os sentimentos evocados pelas imagens. Todas as estudantes foram instruídas a relatar sentimentos de paz, calma e tranquilidade, mas apenas a metade foi exposta a imagens que provavelmente evocariam tais sentimentos, como jardins floridos e paisagens; a outra metade assistiu a vídeos com montagens de cenas cirúrgicas de amputação e queimaduras. Acredita-se que as últimas muito provavelmente estariam mentindo ao relatar seus sentimentos. Os autores buscaram assegurar que essas participantes estivessem motivadas na tarefa informando-as que a capacidade de esconder sentimentos negativos é uma habilidade importante na carreira de uma profissional da Enfermagem. Todas as estudantes foram filmadas enquanto descreviam seus sentimentos, e as fitas foram posteriormente mostradas a observadores que deveriam julgar se cada uma delas estava mentindo ou falando a verdade. Nesse estudo foram usados vários grupos de observadores: estudantes universitários, examinadores de polígrafo, psiquiatras, e membros de diversas forças policiais dos Estados Unidos. Com exceção dos agentes do Serviço Secreto, todos os grupos obtiveram um desempenho de detecção de mentiras em torno de chance. Trabalhos anteriores (DePaulo & Pfeifer, 1986; Kraut & Poe, 1980) já haviam mostrado que profissionais cujo ofício exige reconhecimento de mentiras não obtiveram desempenho superior a leigos em tarefas desse tipo.

Existem várias sugestões de explicação para a baixa acurácia dos detectores nesse tipo de tarefa. Uma delas, e talvez a mais séria, é que pode não haver mudanças observáveis e regulares no comportamento dos emissores enquanto mentem. Não havendo sinais de mentira, não haveria o que se detectar. Nem todos os trabalhos da área obtêm medidas independentes do comportamento dos emissores e, entre aqueles que o fizeram, alguns de fato não encontraram diferenças entre o comportamento dos mentirosos e dos honestos (DePaulo & DePaulo, 1989; Kraut & Poe, 1980), enquanto outros sim (Frank & Ekman, 1997; Mehrabian, 1971).

Uma outra possibilidade está ligada ao desenvolvimento de um repertório de detecção de mentiras no nosso ambiente natural. Ainda que ouvir mentiras, assim como levantar suspeitas sobre a veracidade da fala de outras pessoas, seja fato frequente na vida de qualquer um, não tão frequentemente recebemos qualquer tipo de confirmação para nossas suspeitas. Um motorista pode suspeitar do pedinte no semáforo quando este pede dinheiro para comprar comida (quando na verdade poderá comprar bebida), mas provavelmente o motorista jamais saberá se sua suspeita está certa. Em outras palavras, o ambiente natural não fornece *feedback* adequado para o desenvolvimento de um repertório de detecção de mentiras consistente (DePaulo & Pfeifer, 1986). Algo semelhante pode acontecer com profissionais para os quais reconhecer mentiras é importante (como investigadores criminais); muitas vezes não têm oportunidade de pôr à prova suas suspeitas, o que pode, pelo menos em parte, explicar o baixo desempenho desse tipo de população em tarefas de detecção (DePaulo & Pfeifer, 1986; Kraut & Poe, 1980).

Uma maneira de suplementar essa deficiência do ambiente natural seria configurar em laboratório contingências de treino para a habilidade de detectar mentiras, fornecendo *feedback* para os julgamentos dos detectores. Em seu estudo, Zuckerman, Koestner & Alton (1984) tentaram aumentar a acurácia de detectores na tarefa de detecção de mentiras fornecendo informações sobre as mensagens de vários emissores. Os autores elaboraram um Teste de Detecção de Mentiras que consistia em fitas de vídeo contendo oito mensagens (descrições honestas ou enganosas de pessoas que gostavam ou não, dominavam ou eram submissas), contadas por oito emissores (totalizando 64 mensagens) e apresentadas em dois blocos.

Os detectores foram divididos em cinco grupos, de acordo com o tipo e a quantidade de informação recebida: grupo de controle, que não recebeu qualquer tipo de informação sobre as falas dos emissores; grupo quatro-depois, que recebeu *feedback* depois de cada julgamento feito no primeiro bloco; grupo oito-depois, que recebeu *feedback* para todos os julgamentos nos dois blocos; grupo quatro-antes, que no primeiro bloco recebia informações sobre a veracidade das mensagens *antes* de assisti-las; e grupo misto, que recebia informações pré-mensagem no primeiro bloco e *feedback* no segundo bloco.

Os resultados de Zuckerman e cols. (1984) revelaram que, quanto maior o número de informações dadas aos detectores, maior o aumento na detecção de mentiras do bloco 1 para o bloco 2, sendo o grupo oito-depois aquele que mostrou uma melhora mais acentuada. A melhora do grupo quatro-depois não foi significativamente diferente daquela observada no

grupo de controle. Além disso, o aumento na acurácia não se estendeu à detecção de mentiras de novos emissores. Os autores concluíram, portanto, que uma quantidade maior de *feedback* é necessária para que o treino tenha um efeito mais nítido, e que este efeito não se generaliza entre emissores diferentes.

Diversos trabalhos investigaram o efeito do treino sobre a habilidade de detectar mentiras, e quase todos eles revelaram melhora de desempenho, ainda que com ganhos pequenos, possivelmente devido à brevidade da exposição à situação de treino (ver Frank & Feeley, 2003; Vrij, 2005). Além disso, diferenças individuais nos sinais de mentira já foram apontadas por muitos outros estudos da área (e.g., Vrij, 2005). Devido a elas, o treino de detecção das mentiras de um indivíduo não necessariamente resultará em melhor desempenho na detecção de mentiras de outros indivíduos.

O presente trabalho buscou avaliar o efeito do treino com *feedback* sobre a habilidade de detectar mentiras em uma situação de exposição a um número mais extenso de *feedbacks* do que aquele comumente usado na literatura e em uma situação de interação frente-a-frente entre detector e um número menor de emissores. Restringir a interação a pares fixos (detector-emissor) permite maior controle de diferenças individuais, tanto do repertório prévio do detector quanto da variabilidade dos sinais do emissor. Em um segundo momento, os detectores foram testados com emissores diferentes para verificar se houve generalização da habilidade treinada. Além disso, a aplicação de um treino prolongado possibilitou uma análise passo a passo da aquisição da habilidade de detecção de mentiras.

Método

Participantes

Participaram do estudo 20 estudantes universitários, com idades entre 19 e 45 anos, sendo três do sexo masculino e 17 do sexo feminino. Apesar da assimetria de gênero na composição dos participantes, a literatura tem mostrado que homens e mulheres não diferem em detecção de mentiras

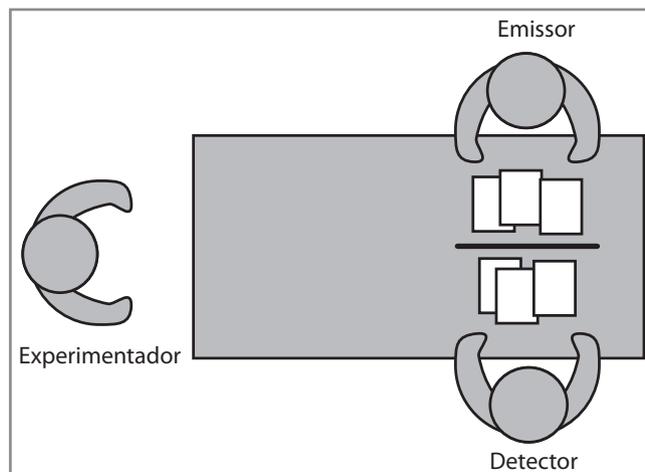


Figura 1. Disposição dos participantes na situação experimental

(Ekman & O'Sullivan, 1991; Vrij, 2005). A participação foi voluntária. Dependendo do papel que desempenhassem no estudo e de seu desempenho na tarefa, os participantes poderiam ganhar R\$ 10,00, R\$ 30,00 ou R\$ 50,00 ao final do experimento. Dos participantes, cinco desempenharam o papel de detector e 15, o de emissor.

Material e cenário

As sessões foram realizadas em uma sala contendo uma mesa com dimensões 150 cm x 100 cm e três cadeiras. A Figura 1 mostra a disposição dos participantes e do experimentador na situação experimental.

Foram utilizadas duas pastas de papel com elástico, uma com o material do detector e outra com o material do emissor. A pasta do emissor continha suas instruções, uma história em quadrinhos completa, uma Folha de Registro e um lápis. A pasta do detector continha suas instruções, as duas primeiras páginas da mesma história em quadrinhos que o emissor havia recebido, e uma Folha de Perguntas (com 10 perguntas preestabelecidas sobre a história). O conteúdo das histórias utilizadas, tal como as perguntas, era geralmente simples e com um mínimo de ambiguidades. Para que um participante não tivesse acesso visual ao material do outro, foi colocada uma divisória de papel-cartão medindo 51 cm x 20 cm entre os dois. Foram utilizados também lápis, cronômetro e marcadores de acrílico para representar a pontuação dos participantes.

Procedimento

A tarefa envolvia a interação entre dois participantes, aqui chamados de *detector* e *emissor*. No início das sessões, cada um recebia a pasta contendo seu material. Na primeira sessão, a leitura das instruções era obrigatória; nas sessões subsequentes, era opcional. Em todas as sessões os participantes deveriam ler a história em quadrinhos (completa, no caso do emissor, e as duas primeiras páginas, no caso do detector). A primeira tentativa começava assim que ambos participantes houvessem terminado de ler seu material.

Em cada tentativa, o detector fazia ao emissor a pergunta da Folha de Perguntas cujo número correspondia àquela tentativa. O emissor deveria, então, responder àquela pergunta, mas sua resposta poderia ser a verdade ou poderia ser uma mentira. A Folha de Registro do emissor indicava se ele deveria falar a verdade ou mentira em cada tentativa, mas não indicava o conteúdo de sua resposta. Assim que o emissor respondia, o detector deveria fazer um julgamento da resposta do emissor, ou seja, deveria dizer a ele “é verdade” ou “é mentira”.

Nas sessões *Sem Feedback* (SF), a tentativa terminava nesse ponto. Nas sessões *Com Feedback* (CF), depois do julgamento do detector, o experimentador dizia a resposta correta (“é verdade” ou “é mentira”) aos participantes, e atribuía pontos a eles conforme mostra o Quadro 1. Marcadores de acrílico eram distribuídos para representar a pontuação, e os participantes tinham acesso visual à própria pontuação e à do outro.

Quadro 1. Distribuição de pontos para o detector e para o emissor por tentativa.

	Emissor diz a verdade		Emissor diz uma mentira	
	Pontuação Detector	Pontuação Emissor	Pontuação Detector	Pontuação Emissor
Detector diz “é verdade”	+1	-	-1	+2
Detector diz “é mentira”	-1	-	+1	-1

No intuito de simular contingências do ambiente real no controle do comportamento do emissor, este só enfrentava riscos (possibilidade de ganhar ou perder pontos) quando estava mentindo. Prevendo ainda um aumento progressivo da acurácia do detector ao longo das sessões, o que resultaria em uma perda cada vez maior de pontos do emissor, programou-se que este ganharia dois pontos caso mentisse com sucesso (ou seja, sem que sua mentira fosse detectada), mas perderia apenas um ponto caso o detector respondesse “mentira” após uma informação de tal tipo do emissor. Essa diferença seria análoga a situações cotidianas nas quais os ganhos do emissor são de maior magnitude quando sua mentira não é detectada.

Além disso, o emissor perdia dois pontos se sua resposta não correspondesse ao indicado naquela tentativa (verdade ou mentira), se a resposta fosse muito ambígua ou não se relacionasse com a pergunta, ou se não respondesse dentro de 30 segundos. Da mesma forma, o detector perderia dois pontos se não fizesse o julgamento em 20 segundos. Em qualquer um desses casos, a tentativa era desconsiderada (e não era contabilizada nos dados), e passava-se para a próxima. A sessão terminava depois de 10 tentativas.

Segue um exemplo de tentativa:

- Detector lê a pergunta em sua Folha de Perguntas: *O que aconteceu quando o vaso de flores caiu no chão?*

- Emissor confere sua Folha de Registro e observa que deve mentir naquela tentativa: *O gato passou por cima dos cacos do vaso e machucou o pé.*

- Detector faz seu julgamento: *É mentira.*

- Experimentador fala a resposta correta (*feedback*): *É mentira.*

- Experimentador dá um ponto ao detector e retira um ponto do emissor.

Cada detector (A, B, C, D e E) realizou um total de 18 sessões, interagindo com três emissores (aqui chamados de E1, e2 e e3). O experimento foi dividido em duas fases. Na Fase I, que compreendeu as 10 primeiras sessões, cada detector interagiu com um único emissor (E1), sendo que nas duas primeiras sessões e na décima, o *feedback* não era dado e não havia pontuação (condição SF). Nas demais, o *feedback* era dado para cada tentativa (condição CF). A Fase II testou a generalização do desempenho do detector para outros emissores. Nas primeiras quatro sessões dessa fase, o detector interagiu com e2, e nas quatro últimas, com e3. As duas primeiras sessões com e1 e e2 eram feitas na condição SF, e as duas seguintes, na condição CF.

O número de respostas verdadeiras (V) programadas na Folha de Registro do emissor variava de dois a oito, sendo que no total das sessões para um mesmo emissor, o número

de respostas verdadeiras programadas era igual ao número de mentiras (M) programadas. Esse balanceamento foi feito separadamente para o conjunto de sessões de cada condição e de cada fase. A ordem das histórias usadas em cada uma das 18 sessões foi escolhida aleatoriamente, e não foi a mesma para diferentes detectores.

De três a quatro sessões poderiam ser feitas em um único dia, consecutivamente. Na última sessão realizada com cada emissor, este e o detector deveriam responder um questionário contendo perguntas abertas relativas à tarefa.

Ao final do experimento, o detector que fez mais pontos (dentre os outros detectores) ganhou um prêmio de R\$ 50,00. O mesmo aconteceu com os emissores da Fase I e da Fase II, mas com prêmios de R\$ 30,00 e R\$ 10,00 respectivamente.

Resultados

A principal medida de desempenho utilizada neste estudo foi a porcentagem de julgamentos corretos (“é verdade” ou “é mentira”) do detector.

A Figura 2 mostra a porcentagem de acertos dos cinco detectores nas 18 sessões, cada um interagindo com três emissores diferentes (os dados da interação do detector A com e2 não foram utilizados pelo fato de o número de tentativas desconsideradas ter sido maior que o número de tentativas válidas).

Pode-se observar, para todos os detectores, elevações e quedas assistemáticas na porcentagem de acertos. Todos eles atingiram porcentagens de acertos iguais ou superiores a 80 em pelo menos uma sessão da Fase I. Os detectores A e B apresentaram porcentagens altas (80% e 89%, respectivamente) já na primeira exposição à condição SF, enquanto que os demais só atingiram porcentagens de acerto acima do acaso (considerado aqui como a faixa de 40% a 60%) nas sessões posteriores dessa fase. Com exceção do detector D, todos os detectores tiveram um aumento na porcentagem de acertos já na primeira sessão da condição CF. Na última sessão da Fase I (décima sessão – condição SF), todos os detectores mantiveram a porcentagem de acertos constante ou em ascensão em relação à sessão anterior (nona sessão – condição CF), com exceção do detector C, que apresentou uma queda. A comparação entre as condições SF e CF indica que, apesar da assistemática dos dados, as maiores porcentagens de acerto foram obtidas na condição CF, com exceção do detector D (para esse detector, a maior porcentagem de acerto ocorreu na segunda exposição à condição SF).

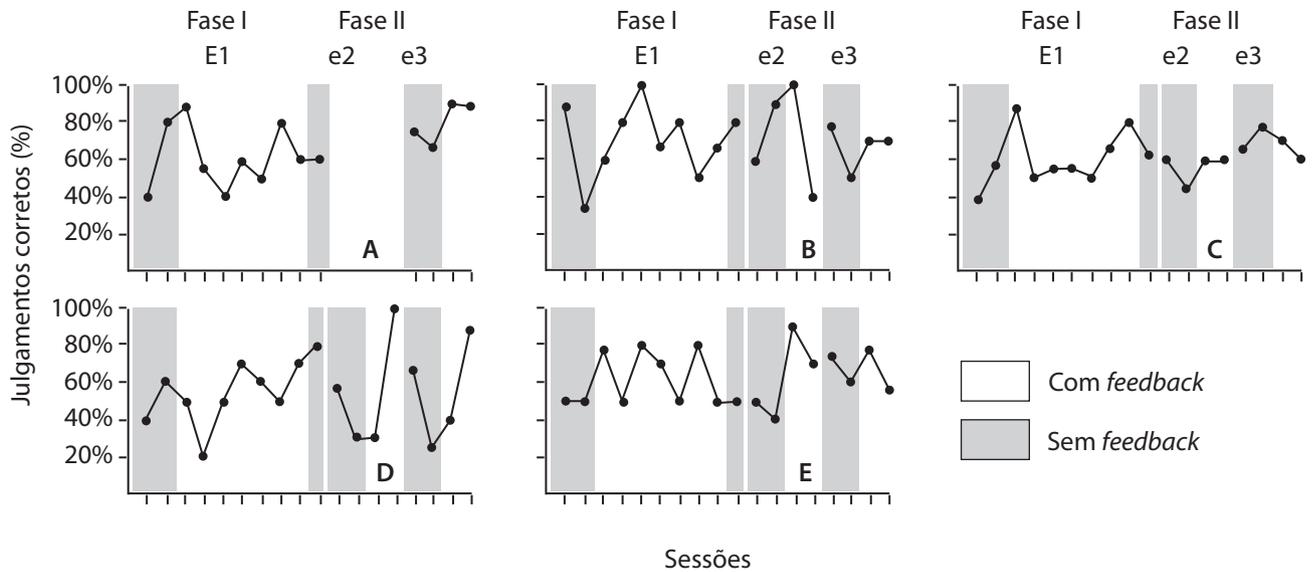


Figura 2. Porcentagem de julgamentos corretos, por sessão, para todos os detectores. A quebra entre as curvas indica as sessões que foram feitas com E1 (Fase I), e2 e e3 (Fase II).

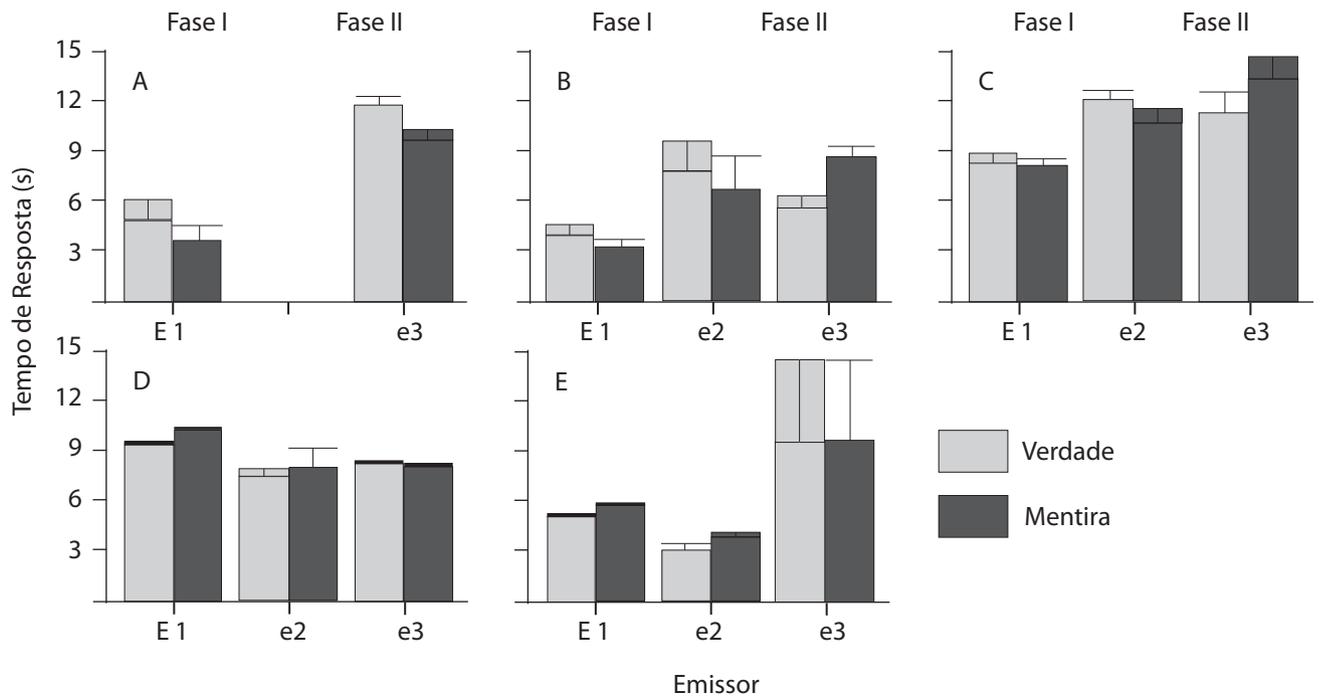


Figura 3. Tempo médio de resposta (TR) de acordo com o tipo (verdade ou mentira) do emissor (colunas verticais) e de acordo com o julgamento (verdade ou mentira) do detector (barras horizontais).

A porcentagem de acertos de todos os detectores na Fase II atingiu valores iguais ou superiores a 75, em pelo menos uma sessão com cada emissor (e2 e e3), com exceção da interação do detector C com e2. Nessa fase, as porcentagens mais altas de acerto nem sempre aconteceram na condição CF.

Foi feito também o registro de duas medidas comportamentais do emissor: o tempo de resposta (TR - tempo transcorrido do momento em que o detector terminava de fazer uma pergunta e o emissor terminava de respondê-la, ou seja, a soma da latência e da duração

da resposta) e a porcentagem de tentativas em que o emissor consultava a história em quadrinhos enquanto dava a resposta.

A Figura 3 mostra os TRs médios das falas dos emissores quando contavam verdades e mentiras, representados pelas colunas verticais. Essa figura mostra também os TRs médios das falas dos emissores que antecediam os julgamentos (verdade ou mentira) do detector, representado pelas barras paralelas às colunas. Em outras palavras, a segunda medida indica quanto tempo, em média, o emissor havia levado para responder nas

tentativas em que o detector fez um julgamento “é verdade” ou “é mentira”. Essa medida permite avaliar se o julgamento do detector foi influenciado pelo TR do emissor.

Para oito dos 14 emissores, o TR para verdade foi maior que para as mentiras, e para os outros seis o contrário foi verdadeiro. A maior diferença observada foi com e3 do detector E, que levou em média 4.9 s a mais para contar verdades do que mentiras. Foram conduzidas análises estatísticas de comparação de médias, sendo utilizado o teste não paramétrico Mann-Whitney, uma vez que a amostra não apresenta distribuição normal. Os resultados indicaram que as diferenças foram significativas, com $p < .05$, apenas para e2 do detector E ($U = 119.5, p = .027$), e com $p < .10$ para E1 do detector A ($U = 953.0, p = .075$) e E1 do detector E ($U = 988.0, p = .094$). Entretanto, apesar da significância estatística, essas diferenças foram pequenas (1.1, 2.5 e 0.7 segundos respectivamente), possivelmente não discrimináveis por um detector humano.

Em relação aos julgamentos, em nove dos 14 pares emissor-detector, o detector, em média, julgou as respostas mais demoradas do emissor como mentira. Os resultados do Mann-Whitney apontam que essas diferenças foram significativas ($p < .10$) para apenas três interações: e3-B ($U = 118.0, p = .042$), e2-D ($U = 89.5, p = .074$) e e3-E ($U = 100.5, p = .053$), o que pode indicar que apenas para esses pares o TR funcionou como dica para os julgamentos do detector (mesmo que não fosse uma dica efetiva).

A Figura 4 mostra a porcentagem de tentativas nas quais os emissores consultavam a história em quadrinhos enquanto respondiam (sendo verdade ou mentira), representada pelas colunas. As barras paralelas às colunas representam a porcentagem de tentativas com consultas seguidas por julgamentos “é verdade” ou “é mentira”. Essa última medida indica se os

julgamentos do detector foram afetados pela ocorrência ou não de consulta por parte do emissor.

Noves emissores consultaram mais vezes a história em quadrinhos quando contavam verdades. A maior diferença observada foi com e3 do detector A (38.9%). Testes qui-quadrado de independência mostraram que a diferença na porcentagem de consultas entre tentativas com verdade e com mentira foi significativa ($p < .10$) para E1-A [$\chi^2 = 11.6 (1, N = 98, p = .001)$], e3-A [$\chi^2 = 6.4 (1, N = 36, p = .011)$], E1-B [$\chi^2 = 5.1 (1, N = 94, p = .024)$] e e2-B [$\chi^2 = 3.2 (1, N = 38, p = .074)$].

Com relação aos julgamentos, em seis dos 14 pares emissor-detector, o detector julgou as respostas com um maior número de consultas como verdades, e em oito pares o oposto ocorreu, ou seja, respostas com um maior número de consultas foram julgadas como mentiras. Essas diferenças foram estatisticamente significativas ($p < .10$) apenas nos pares e3-A [$\chi^2 = 9.2 (1, N = 36, p = .002)$] e E1-D [$\chi^2 = 3.2 (1, N = 99, p = .072)$], o que sugere que apenas nessas interações as consultas funcionaram como dicas para os julgamentos do detector (no primeiro caso, acuradamente).

Ainda que os dados sugiram que, para a maioria dos emissores, nem o TR e nem a porcentagem de consultas funcionaram como sinal de mentira (ou de verdade), e nenhuma das duas medidas comportamentais foi usada como dica de detecção, as respostas dadas nos questionários apontam em uma direção diferente. Quatro emissores descreveram que demoravam mais para mentir ou para falar a verdade, e apenas um deles (e1-E) teve uma diferença no TR estatisticamente significativa. Sete detectores descreveram usar o TR como dica para saber se o emissor estava mentindo, mas de acordo com os testes conduzidos, apenas em três interações o TR diferenciou julgamentos de verdade e de mentira. Em alguns casos, o relato foi oposto ao dado observado, como

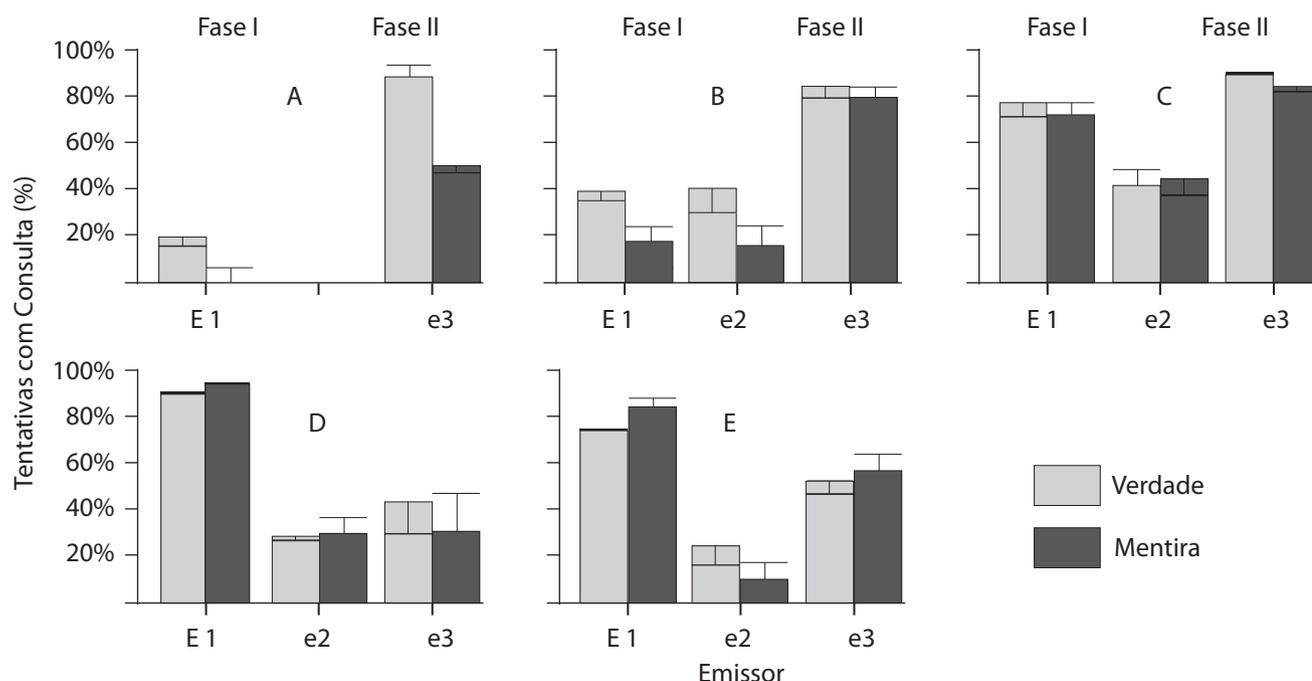


Figura 4. Porcentagem de tentativas com consulta de acordo com o tipo de fala (verdade ou mentira) do emissor (colunas verticais) e de acordo com o julgamento (verdade ou mentira) do detector (barras horizontais).

por exemplo, “a mentira não vinha rapidamente em minha mente. Então demorava a responder” (E1-B).

Discussão

Os dados para as porcentagens de acerto dos detectores mostram que, pelo menos em alguns momentos, a detecção de mentiras de fato ocorreu, ainda que não haja efeitos sistemáticos aparentes da exposição ao *feedback* no desempenho, não havendo também indícios claros de generalização da aprendizagem. Além disso, essas porcentagens foram muito superiores àquelas geralmente apontadas em outros estudos. Em uma revisão da literatura sobre treinamento em detecção de mentiras, Frank e Feeley (2003) fizeram uma meta-análise de 11 estudos dessa natureza, comparando a porcentagem de detecção agregada em grupos de controle com os grupos nos quais houve treino. Esses autores concluíram que, em geral, os ganhos com o treino foram pequenos, ainda que confiáveis (4%, em média). O estudo no qual foi obtida a média mais elevada de porcentagem de acertos foi o de Zuckerman e cols. (1984), no qual essa média correspondeu a 70% para os grupos oito-depois e misto.

Frank e Feeley (2003) acreditam que tais resultados subestimam a potencialidade do treino pelo fato de que a maioria dos estudos não se preocupou em criar situações experimentais de alto-risco e suficientemente similares a situações reais para garantir sua relevância. De acordo com os autores, a falha em atingir esses critérios pode resultar no não aparecimento de sinais de mentira, de forma a dificultar ou impossibilitar qualquer tipo de treino.

O presente estudo criou uma situação experimental que não envolvia mentiras sobre algo relevante para os participantes, e nem envolvia riscos muito altos para o emissor caso fosse pego, aspectos geralmente tratados como fundamentais na literatura (Ekman & O’Sullivan, 1991; Frank & Ekman, 1997; Frank & Feeley, 2003; Iacono & Lykken, 2002). Além disso, ao contrário do estudo de Zuckerman e cols. (1984), que selecionou os emissores com base em altas porcentagens de detecção obtidas previamente, não houve nenhuma preocupação em garantir que os emissores se comportariam diferencialmente enquanto mentiam. Ainda assim, no presente estudo foram obtidas porcentagens de acertos de até 100% em uma única sessão, e todos os detectores em algum momento chegaram a um nível de acertos de pelo menos 85%.

Por que isso aconteceu? A primeira, e mais óbvia explicação, diz respeito ao tratamento dos dados. Todos os estudos aos quais se fez referência até agora trabalharam com médias de grupo, de forma que valores altos na porcentagem de acertos são ofuscados por valores baixos, enquanto o presente estudo analisou os dados individualmente. Uma segunda explicação, e talvez a mais importante, diz respeito a diferenças no procedimento. Enquanto no trabalho de Zuckerman e cols. (1984) os detectores recebiam no máximo oito *feedbacks* por remetente, no presente estudo recebiam até 70 *feedbacks* para um único emissor na Fase I, e até 20 por emissor na Fase II. Outra diferença possivelmente relevante foi o fato de que, neste trabalho, detector e emissor interagiram frente a frente, e não mediante filmagem. Encarar o detector (que faz ao mesmo tempo papel de interrogador) de frente pode

tornar mais difícil para o emissor mentir com sucesso. Além disso, nas fitas de vídeo, os detectores dos outros estudos não tinham acesso a dicas importantes como o tempo de resposta (que nos vídeos era padronizado). Dados sobre efeitos de interação frente a frente na porcentagem de detecção na literatura, entretanto, mostram resultados conflitantes (Feeley & Young, 1998).

As altas porcentagens de acerto permitem sugerir, portanto, que, em alguns momentos, os detectores estavam discriminando diferenças no comportamento do emissor quando ele mentia a quando era honesto. Contudo, conclusões que vão para além disso merecem algumas ressalvas.

Primeiro: detectar mentiras foi uma habilidade treinada ou pré-existente no repertório dos detectores? A linha de base utilizada neste estudo (duas primeiras sessões da condição SF) pode ter sido muito pequena para ser representativa do repertório pré-experimental dos detectores, principalmente porque houve muitas oscilações no desempenho. Entretanto, o fato de que para todos eles (com exceção do detector D) a porcentagem de acertos aumentou logo na primeira sessão com *feedback* é um indicativo de que o treino teve efeito. O mesmo acontece na Fase II do experimento: quase sempre a inserção do *feedback* foi acompanhada por um aumento na porcentagem de acertos.

Segundo: a aprendizagem foi favorecida mais pelo treino (fornecimento de *feedback*) ou simplesmente pela prática (exposição a falas verdadeiras e mentirosas do emissor)? No trabalho de Zuckerman e cols. (1984), o grupo de controle apresentou um aumento significativo na porcentagem de acertos, e foi o único para o qual este aumento também foi observado entre diferentes emissores (generalização). No presente trabalho, pode-se observar, em alguns casos, aumentos grandes na porcentagem de acertos da primeira para a segunda sessão (antes da inserção do *feedback*), o que fortalece a ideia de que a prática sozinha pode melhorar o desempenho dos detectores. O delineamento deste estudo não pode dar uma resposta satisfatória a esta questão: seria necessário colocar detectores em exposição prolongada ao procedimento sem, em nenhum momento, fornecer-lhes *feedback*.

Terceiro: a habilidade adquirida é consistente? A curva de desempenho de todos os detectores mostra grandes oscilações, o que pode significar duas coisas: uma variação normal e esperada nas fases iniciais da aquisição de um repertório novo qualquer; ou variações no comportamento do emissor. Ainda que o primeiro fator possa ter de fato contribuído para as oscilações no desempenho do detector, se fosse este o fator preponderante esperar-se-ia algum nível de estabilização nas últimas sessões (pelo menos para um mesmo emissor), o que não aconteceu. Por outro lado, há vários indícios de variação no comportamento do emissor. Afinal, ele também tinha acesso ao *feedback* dado pelo experimentador ao detector, e sabia quando o detector acertava e quando ele errava. Da mesma forma que o detector observava o comportamento do emissor, escolhia sinais para julgar se a fala do emissor era verdade ou mentira, e depois confirmava ou não suas suspeitas com o *feedback*, o emissor poderia estar observando o próprio comportamento e suspeitando de coisas que o detector poderia estar usando como dicas para fazer seu julgamento; o *feedback* confirmaria ou não suas suspeitas.

Dessa forma, as quedas no desempenho poderiam representar o desaparecimento de sinais que estavam funcionando como dica para o detector (pela aprendizagem do emissor), ao passo que as elevações posteriores poderiam representar a aprendizagem do detector para discriminar novos sinais que o emissor ainda não aprendeu a disfarçar ou omitir. Pode ter ocorrido, portanto, uma aprendizagem por parte do emissor, que passava a “mentir melhor” e dificultava a tarefa do detector. Essas considerações enfraquecem uma posição de que essas oscilações foram aleatórias.

O seguinte relato (detector E em relação ao e3) parece corroborar essa ideia: “*Em algumas ocasiões, no início percebi que ela ficava séria quando dizia a verdade, e sorria quando mentia. Mas parece ter percebido isto também, e passou a fazer o contrário, tornando mais difícil identificar*”. Outro exemplo: “*No início, percebi que a maioria das respostas curtas eram mentiras. Depois as mentiras passaram a ser mais elaboradas, o que dificultou mais a identificação*” (detector E sobre E1). Além disso, vários emissores relataram consultar a história para disfarçar quando mentiam, uma vez que consultas deveriam ser a princípio mais importantes (e frequentes) nas tentativas que o emissor deveria dizer o que de fato aconteceu na história (verdade). Feeley e Young (1998) não acreditam, entretanto, que emissores tenham as informações necessárias sobre seus próprios correlatos comportamentais da mentira, e mesmo quando têm, não são capazes de utilizar essas informações para adaptar seu comportamento de acordo (Vrij, Semin, & Bull, 1996).

É claro que a aprendizagem de detectores e emissores pode ter estado sujeita a outras nuances, e toda essa dinâmica pode ter operado sem que algumas vezes os participantes tivessem consciência (ou, pelo menos, consciência parcial) do processo. A Análise do Comportamento tem demonstrado que aumentos no desempenho muitas vezes não são acompanhados de relatos das contingências responsáveis pela seleção daquele desempenho. Em alguns momentos, o desempenho aumenta, mas só depois “emerge” o relato (ou a consciência) dessas contingências (e.g., Borges, 2002). Comparações entre os dados e os relatos dos participantes sugerem não só que os emissores não identificavam diferenças no seu próprio comportamento nas condições verdade e mentira, mas que os detectores também não souberam apontar as dicas que eles próprios usavam para fazer seus julgamentos, de forma que os relatos feitos nos questionários pareciam mostrar mais descrições estereotipadas do comportamento das pessoas enquanto mentem (demora para responder, nervosismo) do que as mudanças comportamentais relevantes. Outros trabalhos encontraram indicativos de correspondência entre o relato dos detectores e as pistas que eles usavam de fato, mesmo que estes estivessem usando pistas estereotipadas e inaccuradas na detecção (Anderson, DePaulo, Ansfield, Tickle, & Green, 1999).

Uma última questão é: houve generalização da aprendizagem? Se o detector de fato aprendeu a discriminar sinais de mentira na Fase I, e os emissores da Fase II ainda não passaram por nenhum tipo de aprendizagem, poder-se-ia esperar altas porcentagens de acerto com e2 e e3 já nas primeiras sessões da condição SF, mas isso nem sempre aconteceu. Na verdade, na primeira sessão com cada emissor da Fase II, a porcentagem de acertos nunca foi alta, ainda que sempre

(com exceção de e2 do detector C) tenha chegado a pelo menos 75% nas sessões posteriores. É possível que o treino com o primeiro emissor tenha não só colocado o julgamento do detector sob controle discriminativo de sinais específicos, mas o tornado mais sensível a alterações comportamentais sutis de forma geral. E mesmo que o detector não soubesse exatamente o que procurar em novos emissores para saber se estavam mentindo, poderia discriminar com maior rapidez e precisão os novos sinais, assim como as mudanças que ocorriam à medida que o *feedback* modificava o comportamento do emissor. A complexidade e a variação do comportamento dos emissores da Fase I, assim como importantes diferenças individuais entre diferentes emissores, fazem da generalização direta do efeito do treino da Fase I para a Fase II pouco provável.

Conclusão

De forma geral, os dados do presente estudo mostram que, em algum grau, os participantes foram capazes de detectar mentiras, mostrando um desempenho acima do acaso, diferente daquele geralmente apontado na literatura (DePaulo, Kirkendol, Tang & O'Brien, 1988). O delineamento utilizado não permitiu responder de forma clara sobre os efeitos do treino com *feedback*, nem sobre a generalização da habilidade treinada. Ainda assim, o presente trabalho oferece uma contribuição metodológica à área, isto é, um procedimento de treino que permite um número muito grande de tentativas de detecção e, conseqüentemente, de conseqüenciação do julgamento do detector na interação com um único emissor. O conteúdo das mentiras em outros estudos (como simulações de roubo, expressão de opiniões fortes) não possibilitava um número muito grande de mensagens para um mesmo emissor sem que suas falas ficassem problematicamente repetitivas (DePaulo & cols., 1988; Frank & Ekman, 1997; Kraut & Poe, 1980), de forma que se torna necessário usar vários emissores para falar de uma mesma situação. Trabalhos posteriores devem buscar controlar variações no comportamento do emissor por meio da restrição do acesso a informações que promovam tais variações (como o *feedback*), assim como examinar mais detalhadamente as mudanças nos comportamentos e estratégias de detector e emissor ao longo desse processo de aprendizagem que é, essencialmente, dinâmico e interacionista.

Referências

- Anderson, D. E., DePaulo, B. M., Ansfield, M. E., Tickle, J. J., & Green, E. (1999). Beliefs about cues to deception: Mindless stereotypes or untapped wisdom? *Journal of Nonverbal Behavior*, 23, 67-89.
- Borges, F. S. (2002). *Comportamento de escolha em humanos: uma diferença entre fazer e dizer*. Dissertação de Mestrado, Universidade Católica de Goiás, Goiânia.
- DePaulo, P. J., & DePaulo, B. M. (1989). Can deception by salespersons and customers be detected through nonverbal behavioral cues? *Journal of Applied Social Psychology*, 19, 1552-1577.

- DePaulo, B. M., & Kashy, D. A. (1998). Everyday lies in close and casual relationships. *Journal of Personality and Social Psychology*, 74, 63-79.
- DePaulo, B. M., Kashy, D. A., Kirkendol, S. E., Wyer, M. M., & Epstein, J. A. (1996). Lying in everyday life. *Journal of Personality and Social Psychology*, 70, 979-995.
- DePaulo, B. M., Kirkendol, S. E., Tang, J., & O'Brien, T. P. (1988). The motivational impairment effect in the communication of deception: Replications and extensions. *Journal of Nonverbal Behavior*, 12, 177-202.
- DePaulo, B. M., & Pfeifer, R. L. (1986). On-the-job experience and skill at detecting deception. *Journal of Applied Social Psychology*, 16, 249-267.
- Ekman, P. (2001). *Telling lies: Clues to deceit in the marketplace, politics and marriage* (3ª ed.). Nova Iorque: Norton.
- Ekman, P., & O'Sullivan, M. (1991). Who can catch a liar? *American Psychologist*, 46, 913-920.
- Feeley, T. H., & Young, M. J. (1998). Humans as lie detectors: Some more second thoughts. *Communication Quarterly*, 46, 109-126.
- Frank, M. G., & Ekman, P. (1997). The ability to detect deceit generalizes across different types of high-stake lies. *Journal of Personality and Social Psychology*, 72, 1429-1439.
- Frank, M. G., & Feeley, T. H. (2003). To catch a liar: Challenges for research in lie detection training. *Journal of Applied Communication Research*, 31, 58-75.
- Iacono, W. G., & Lykken, D. T. (2002). Polygraph tests. Em D. L. Faigman, D. H. Kaye, M. J. Saks & J. Sanders (Orgs.), *Modern scientific evidence: The law and science of expert testimony*, Vol. 2 (pp. 427-538). St. Paul: West Publishing.
- Kraut, R. E., & Poe, D. (1980). Behavioral roots of person perception: The deception judgments of customs inspectors and laymen. *Journal of Personality and Social Psychology*, 39, 784-798.
- Mehrabian, A. (1971). Nonverbal betrayal of feeling. *Journal of Experimental Research in Personality*, 5, 64-73.
- Vrij, A. (2005). *Detecting lies and deceit* (5ª ed.). Chinchester: John Willey.
- Vrij, A., Semin, G. R., & Bull, R. (1996). Insight into behavior displayed during deception. *Human Communication Research*, 22, 544-562.
- Zuckerman, M., Koestner, R. E., & Alton, A. (1984). Learning to detect deception. *Journal of Personality and Social Psychology*, 46, 519-528.

Recebido em 01.10.07
Primeira decisão editorial em 01.09.08
Versão final em 21.09.08
Aceito em 20.11.08 ■